

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-323340

(43)Date of publication of application : 14.11.2003

(51)Int.Cl.

G06F 12/08

G06F 3/06

G06F 12/12

(21)Application number : 2002-129748

(71)Applicant : HITACHI LTD

(22)Date of filing : 01.05.2002

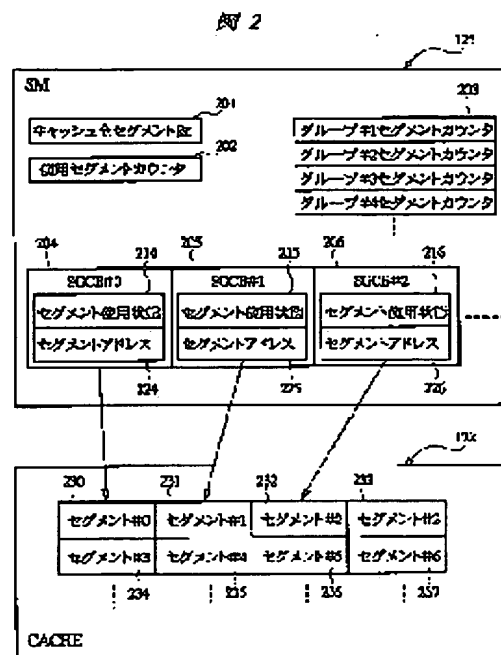
(72)Inventor : NAKANISHI HIROAKI  
TAKEUCHI HISAHARU  
KUROKAWA ISAMU  
KAWAGUCHI KATSUHIRO

(54) CONTROL METHOD OF STORAGE SUB-SYSTEM AND STORAGE SUB- SYSTEM

(57)Abstract:

**PROBLEM TO BE SOLVED:** To improve the performance of the whole of a storage sub-system by preventing a deviation of a using quantity of a cache memory between a plurality of disk groups constituting a storage device.

**SOLUTION:** The inside of the common memory 125 of a disk control device 102 is provided with the whole cache segment number 201 for constituting a cache memory part 122, a using segment counter 202 for indicating a cache using quantity of the whole sub-system, and a group unit segment counter 203 for controlling a segment using quantity of a magnetic disk group unit. Performance reduction in the whole of the storage sub-system caused when a cache utilization factor deviates to a specific disk group is prevented by restraining projection of the cache utilization factor of an individual disk group according to the size of the cache utilization factor of the whole sub-system.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the  
examiner's decision of rejection or application  
converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of  
rejection][Date of requesting appeal against examiner's decision  
of rejection]

[Date of extinction of right]



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2003-323340

(P2003-323340A)

(43) 公開日 平成15年11月14日 (2003. 11. 14)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テマコード <sup>*</sup> (参考)
G 0 6 F 12/08	5 2 3	G 0 6 F 12/08	5 2 3 E 5 B 0 0 5
	5 4 3		5 4 3 B 5 B 0 6 5
3/06	3 0 2	3/06	3 0 2 A
12/12	5 5 7	12/12	5 5 7

審査請求 未請求 請求項の数 3 O L (全 8 頁)

(21) 出願番号 特願2002-129748 (P2002-129748)

(22) 出願日 平成14年5月1日 (2002. 5. 1)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 中西 弘晃

神奈川県小田原市中里322番地2号 株式会社日立製作所 R A I D システム事業部内

(72) 発明者 竹内 久治

神奈川県小田原市中里322番地2号 株式会社日立製作所 R A I D システム事業部内

(74) 代理人 100080001

弁理士 筒井 大和

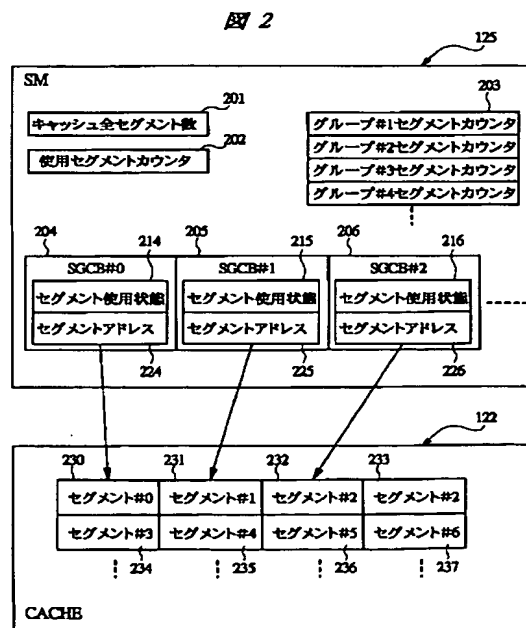
最終頁に続く

(54) 【発明の名称】 記憶サブシステムの制御方法および記憶サブシステム

(57) 【要約】

【課題】 記憶装置を構成する複数のディスクグループ間でのキャッシュメモリの使用量の偏りを防止して、記憶サブシステム全体の性能を向上させる。

【解決手段】 ディスク制御装置102の共有メモリ125内に、キャッシュメモリ部122を構成するキャッシュ全セグメント数201と、サブシステム全体のキャッシュ使用量を示す使用セグメントカウンタ202と、磁気ディスクグループ単位のセグメント使用量を管理するグループ単位セグメントカウンタ203を設け、サブシステム全体のキャッシュ使用率の大小に応じて、個々のディスクグループのキャッシュ使用率の突出を抑止することで、特定のディスクグループにキャッシュ使用率が偏ることに起因する記憶サブシステム全体の性能低下を防止する。



## 【特許請求の範囲】

【請求項 1】 複数の記憶装置と、前記記憶装置と上位装置との間で授受される情報が一時的に格納されるキャッシュメモリを備えた記憶制御装置と、を含む記憶サブシステムの制御方法であって、

前記記憶サブシステム全体の前記キャッシュメモリの使用量を評価する第 1 閾値と、前記記憶装置のグループの各々の前記キャッシュメモリの使用量を評価する第 2 閾値とを設定し、

前記記憶サブシステム全体の前記キャッシュメモリの使用量の大小に基づいて、前記記憶装置のグループ単位の前記キャッシュメモリの使用量を動的に制限することを特徴とする記憶サブシステムの制御方法。

【請求項 2】 請求項 1 記載の記憶サブシステムの制御方法において、前記第 1 閾値として複数の値が多段階に設定され、前記記憶サブシステム全体の前記キャッシュメモリの使用量を多段階に評価して各段階毎に前記記憶装置のグループ単位の前記キャッシュメモリの使用量を動的に制限することを特徴とする記憶サブシステムの制御方法。

【請求項 3】 複数の記憶装置と、前記記憶装置と上位装置との間で授受される情報が一時的に格納されるキャッシュメモリを備えた記憶制御装置と、を含む記憶サブシステムであって、

前記記憶制御装置は、前記記憶サブシステム全体の前記キャッシュメモリの使用量を管理する第 1 記憶手段と、前記記憶装置のグループの各々の前記キャッシュメモリの使用量を管理する第 2 記憶手段と、前記記憶サブシステム全体の前記キャッシュメモリの使用量の大小に基づいて前記記憶装置のグループ単位の前記キャッシュメモリの使用量を動的に制限する制御論理とを含むことを特徴とする記憶サブシステム。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は記憶サブシステムの制御技術および記憶サブシステムに関し、特に、キャッシュメモリ付きの記憶制御装置を備えた記憶サブシステム等に適用して有効な技術に関する。

## 【0002】

【従来の技術】 磁気ディスク制御装置におけるキャッシュメモリは従来 LRU (Least Recently Used) 管理方式によるもっとも利用されているデータをキャッシュメモリ上に蓄えることで性能向上を図ってきた。また特定ディスク単位にキャッシュメモリ上にキャッシュメモリ全体の一部を常に割り当てる機能を提供し、特定ディスクに対する入出力要求の性能向上を行ってきた。

## 【0003】

【発明が解決しようとする課題】 たとえば、ディスクアレイを制御する磁気ディスク制御装置が管理する磁気デ

ィスク数が増加し続けており、サブシステム全体でのキャッシュメモリの効率的な利用を考えていた LRU 管理方式では、特定の磁気ディスクに対するデータの入出力要求が極めて高まった場合のサブシステム内の他ディスクに対する性能への影響が考えられていなかった。

【0004】 このため入出力要求が高くキャッシュメモリ使用量が多くなる磁気ディスクがサブシステム内に含まれていた場合、磁気ディスク制御装置が管理する磁気ディスク数が増加すると他の磁気ディスクに対する入出力要求への処理時間が延びてしまう傾向が出てきた。

【0005】 また特定の磁気ディスクに対して常にキャッシュメモリを一部割り当てて性能向上を図り、他ディスクに影響を与えない機能も提供されてきたが、動的に管理できていないことから一時的な負荷が高まる状況では回避できなかった。

【0006】 本発明の目的は、記憶装置を構成する複数のディスクやディスクグループの各々の間でのキャッシュメモリの使用量の偏りを防止して、記憶サブシステム全体の性能を向上させることにある。

## 【0007】

【課題を解決するための手段】 本発明は、複数の記憶装置と、この記憶装置と上位装置との間で授受される情報が一時的に格納されるキャッシュメモリを備えた記憶制御装置と、を含む記憶サブシステムの制御方法であって、

記憶サブシステム全体のキャッシュメモリの使用量を評価する第 1 閾値と、記憶装置のグループの各々のキャッシュメモリの使用量を評価する第 2 閾値とを設定し、記憶サブシステム全体のキャッシュメモリの使用量の大小に基づいて、記憶装置のグループ単位の前記キャッシュメモリの使用量を動的に制限するものである。

【0008】 より具体的には、一例として、複数の磁気ディスクと、磁気ディスクに対するデータ入出力要求を発行する中央処理装置との間に介在し、複数の磁気ディスクを制御する磁気ディスク制御装置とを含む記憶サブシステムにおいて、磁気ディスク制御装置内に中央処理装置から要求された複数の磁気ディスクへの入出力データを一時的に蓄えるキャッシュメモリを有し、サブシステム全体で使用されるキャッシュメモリ使用量と、任意の複数の磁気ディスクからなる磁気ディスクグループ単位で使用されるキャッシュメモリ使用量を管理し、入出力要求に対してキャッシュメモリを新たに割り当てる際に、サブシステム全体として最大限使用可能なキャッシュメモリ使用量の閾値 H と、磁気ディスクグループ単位の前記キャッシュメモリ使用量の閾値 M を用い、サブシステム全体のキャッシュメモリ使用量と磁気ディスクグループ単位の前記キャッシュメモリ使用量とを比較し、突出した磁気ディスクグループ単位の前記キャッシュメモリ使用量を動的に制限するキャッシュメモリ管理技術を提供する。

【0009】 これにより、従来技術では管理しきれない

なかった複数の磁気ディスクやディスクグループに対するキャッシュメモリ割り当てにおいて、他磁気ディスクやディスクグループに対する影響を考慮に入れたキャッシュメモリ管理技術を実現でき、記憶サブシステム全体の性能向上が可能となる。

【0010】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照しながら詳細に説明する。

【0011】図1は本発明の一実施の形態である記憶サブシステムを含むデータ処理システムの構成を示す図である。

【0012】本実施の形態の記憶サブシステムは、中央処理装置であるCPU101等の上位装置との間におけるデータの授受を制御するディスク制御装置であるDKC102と、ディスク駆動装置であるDKU103と、CPU101とDKC102間を結合するケーブル104と、DKC102とDKU103間を結合するケーブル105で構成される。

【0013】上位装置であるCPU101内の構成としては、オペレーティングシステム等を実行する演算処理部(IP)111と、前記プログラム及び、プログラムが処理するデータ等を格納する主記憶装置部(MS)112と、外部入出力装置とのデータ転送を制御する及び、入出力装置とのインタフェースを制御する入出力チャネル部(CH)113と、これらデータ転送を制御するシステムコントローラ部(SC)114から構成される。

【0014】またDKC102内の構成としては、CPU101のチャネル制御部とインタフェースを制御するポート部(PORT)121と、データを高速処理する為のキャッシュメモリ部(CACHE)122と、DKC102内のデータを管理する制御プロセッサ部123と、これらデータ転送を制御データ転送制御部124と、I/Oの制御やデータを管理する為の制御情報等を格納する共有メモリ(SM)125から構成される。

【0015】またDKU103内には、たとえば磁気ディスク装置等からなり、データを格納する複数のディスク装置131が存在する。複数のディスク装置131は必要に応じてRAID等のディスクアレイを構成することができる。

【0016】図2は共有メモリ(SM)内のキャッシュ管理領域とキャッシュメモリ部(CACHE)との関係を示した図である。

【0017】キャッシュメモリ部122はセグメントという単位で分割され管理されている。この複数のセグメント230、231、232、...の各々を共有メモリ125内の複数のセグメント制御ブロック(SGCB)204、205、206、...にて各セグメントの使用状況を管理する。このセグメントとSGCBは一对で管理される。SGCB204、205、20

6、...の各々にはセグメント使用状態214、215、216、...と、セグメントアドレス224、225、226、...を情報として持ち、対応する個々のセグメントを管理する。

05 【0018】また共有メモリ125内には磁気ディスクグループ単位のセグメント使用量を管理するグループ単位セグメントカウンタ203がある。また、全サブシステムのセグメント使用量を管理する使用セグメントカウンタ202がある。各セグメントカウンタの加減算はセグメントを使用状態にする場合に加算され、セグメント未使用状態にする場合に減算される。各磁気ディスクグループ単位のグループ単位セグメントカウンタ203におけるセグメント使用量の加減算時に使用セグメントカウンタ202も同時に加減算され、サブシステム全体のセグメント使用量と磁気ディスクグループ単位のセグメント使用量の比較を行う場合に使用される。

15 【0019】またキャッシュメモリ部122の全体のセグメント数(実装数)が保持されるキャッシュ全セグメント数201が共有メモリ125にあり、キャッシュ利用率算出に使用される。

20 【0020】図3はキャッシュセグメントの新規確保時の処理の一例を示したフローチャートである。

【0021】制御プロセッサ部123はPORT121がCPU101より受領したデータ転送指示に対してキャッシュメモリ部122上にデータが存在しているか判断し、キャッシュメモリ部122上にデータ転送指示されたデータが無い場合、キャッシュメモリ部122に利用可能な空き領域があるか判断処理を開始する(ステップ301)。

30 【0022】まず、キャッシュ全セグメント数201、サブシステム内の使用セグメントカウンタ202、前記データ転送指示を受けたデータが属する特定グループのグループ単位セグメントカウンタ203を共有メモリ125から読み込み(ステップ302)、サブシステム内キャッシュ利用率((使用セグメントカウンタ202の値)÷(キャッシュ全セグメント数201)×100%)、特定ディスクグループ内キャッシュ利用率((グループ単位セグメントカウンタ203の値)÷(キャッシュ全セグメント数201)×100%)を算出する(ステップ303)。

40 【0023】次にサブシステム内キャッシュ利用率が閾値M%(たとえば60%)以上かを判定し(ステップ304)、M%以上ならば更に閾値H%(たとえば70%)以上かを判定し(ステップ305)、H%以上ならば更に特定グループ内のキャッシュ利用率が閾値L%(たとえば30%)以上かを判定し(ステップ306)、L%未満であればデータ入出力要求の流入制限するために待ち時間T1(たとえば0.05秒)の待ち状態を設定し(ステップ308)、キャッシュ空きセグメント確保後失敗処理(ステップ311)にて指定時間待ち状態

を維持し、待ち時間T1の時間経過後キャッシュ空きセグメント再確保処理へ進む。

【0024】(ステップ306)の判定において、特定グループ内のキャッシュ利用率が閾値L%以上の場合、流入制限として待ち時間T2(たとえば1.0秒)の待ち状態設定し(ステップ309)、キャッシュ空きセグメント確保後失敗論理(ステップ311)へ進む。

【0025】また(ステップ305)の判定において、サブシステム内キャッシュ利用率が閾値H%未満の場合、特定グループ内キャッシュ利用率が閾値L%以上かを判定し(ステップ307)、閾値L%以上の場合、流入制限として待ち時間T2(たとえば1.0秒)の待ち状態設定し(ステップ309)、キャッシュ空きセグメント確保後失敗論理(ステップ311)へ進む。

【0026】また(ステップ307)の判定において特定グループ内キャッシュ利用率が閾値L%未満の場合、流入制限を行わずキャッシュメモリ部122の空きセグメントを確保して割り当て(ステップ310)、キャッシュ空きセグメント確保後論理(ステップ312)へと進む。

【0027】また(ステップ304)の判定においてサブシステム内キャッシュ利用率が閾値M%未満の場合、流入制限を行わずキャッシュメモリ部122の空きセグメントを確保して割り当て(ステップ310)、キャッシュ空きセグメント確保後論理(ステップ312)へと進む。

【0028】図4はキャッシュメモリ部122のセグメントの再確保要求時の処理の一例を示すフローチャートである。

【0029】キャッシュメモリ部122のセグメント新規確保時に流入制限として待ち状態を設定された場合、待ち状態を終了後、セグメント再確保要求される(ステップ401)。

【0030】まずサブシステム内のキャッシュセグメント数、サブシステム内の使用セグメントカウンタ202、グループ単位セグメントカウンタ203を共有メモリ125から読み込み(ステップ402)、サブシステム内キャッシュ利用率、特定グループ単位キャッシュ利用率を算出する(ステップ403)。

【0031】次にサブシステム内キャッシュ利用率がU%(たとえば75%)未満を判定し(ステップ404)、U%未満の場合、キャッシュメモリ部122のセグメントを確保し(ステップ405)、キャッシュ空きセグメント確保後論理(ステップ406)へ進む。

【0032】(ステップ404)の判定時ににおいてサブシステム内キャッシュ利用率がU%以上の場合、待ち時間T3(たとえば0.05秒)を設定し(ステップ407)、待ち状態に入る。待ち時間T3経過を判定し(ステップ408)、待ち時間経過後に(ステップ402)へ進み再度キャッシュ利用状態をチェックし、キャッシ

ュメモリ部122のセグメント利用可能か否かの判断論理へ進みセグメントが確保されるまで繰り返される。

【0033】図5、図6、図7、図8はキャッシュメモリ部122の利用率を表した棒グラフの一例である。

【0034】サブシステム全体のキャッシュ利用率と各ディスクグループ内のキャッシュ利用率を示しており、サブシステム全体のキャッシュ利用率を全体と表し、各ディスクグループを#1、#2、#3、#4であらわしている。

【0035】図5はサブシステム全体のキャッシュ利用率がM%を下回る場合であり、この条件では各ディスクグループにおいてキャッシュメモリ部122のセグメントを無条件に確保可能である。

【0036】図6はサブシステム全体のキャッシュ利用率がM%~H%の間にあり、各ディスクグループ内のキャッシュ利用率はL%を下回る場合であり、この条件では各ディスクグループにおいてキャッシュメモリ部122のセグメントを無条件に確保可能である。

【0037】図7はサブシステム全体のキャッシュ利用率がM%~H%の間にあり、#3ディスクグループ内のキャッシュ利用率のみがL%を上回る場合である。この条件では#1、#2、#4の各ディスクグループにおいてキャッシュメモリ部122のセグメントを無条件に確保可能である。

【0038】しかし#3のディスクグループにおいてはキャッシュメモリ部122のセグメント確保時に待ち時間T2(たとえば1.0秒)の待ち状態を経過したのち再度セグメント確保可能か判断され、サブシステム全体のキャッシュ利用率によりセグメント確保可能か判断される。

【0039】図8はサブシステム全体のキャッシュ利用率がH%を上回り、#3ディスクグループ内のキャッシュ利用率のみがL%を上回る場合である。この条件では#1、#2、#4の各ディスクグループにおいてキャッシュメモリ部122のセグメントを待ち時間T1(たとえば0.05秒)の待ち状態を経過したのち再度セグメント確保可能か判断され、サブシステム全体のキャッシュ利用率によりセグメント確保可能か判断される。#3のディスクグループにおいてはキャッシュメモリ部122のセグメント確保時に待ち時間T2(たとえば1.0秒)の待ち状態を経過したのち再度セグメント確保可能か判断され、サブシステム全体のキャッシュ利用率によりセグメント確保可能か判断される。

【0040】以上説明したように、本実施の形態では、サブシステム全体におけるキャッシュ利用率が既定の閾値を超過した場合に、たとえばキャッシュメモリの割当処理の待ち時間を調整することで、突出したディスクグループ単位のキャッシュ使用量を動的に制限するので、特定のディスクグループにキャッシュメモリの使用量が集中することに起因するシステム全体の性能低下を確実に

に防止できる。

【0041】また、サブシステム全体におけるキャッシュ利用率、および個々のディスクグループのキャッシュ利用率の各々を管理するための閾値を複数段階に設定することで、サブシステム全体および個々のディスクグループにおける負荷の状況に応じてキャッシュ使用量をきめ細かく管理でき、キャッシュメモリの可用性を向上させることが可能になる。

【0042】これにより、従来技術では管理しきれいなかった複数の磁気ディスクやディスクグループに対するキャッシュ割り当てにおいて、他磁気ディスクやディスクグループに対する影響を考慮に入れたキャッシュメモリ管理を実現でき、記憶サブシステム全体の性能向上が可能となる。

【0043】本願の特許請求の範囲に記載された発明を見方を変えて表現すれば以下の通りである。

【0044】複数の磁気ディスクと、複数の磁気ディスクを制御する磁気ディスク制御装置と、磁気ディスクに対するデータ入出力要求を発行する中央処理装置からなるサブシステムにおいて、磁気ディスク制御装置内に中央処理装置から要求された複数の磁気ディスクへ入出力データを一時的に蓄えるキャッシュを有し、サブシステム全体で使用されるキャッシュ使用量と、任意の複数磁気ディスクからなる磁気ディスクグループ単位で使用されるキャッシュ使用量を管理し、入出力要求に対してキャッシュを新たに割り当てる際に、サブシステム全体として最大限使用可能なキャッシュ使用量の閾値Hと、磁気ディスクグループ単位のキャッシュ使用量の閾値Mを用い、サブシステム全体のキャッシュ使用量と磁気ディスクグループ単位のキャッシュ使用量と比較し、突出した磁気ディスクグループ単位のキャッシュ使用量を動的に制限することを特徴とするキャッシュ管理方式。

【0045】以上本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【0046】たとえば、流入制限のための待ち時間T1～T3の値は、上述の実施の形態に例示した値にかぎらず、種々変更可能である。

【0047】

【発明の効果】記憶装置を構成する複数のディスクやディスクグループの各々の間でのキャッシュメモリの使用量の偏りを防止して、記憶サブシステム全体の性能を向上させることができる。

【図面の簡単な説明】

【図1】本発明の一実施の形態である記憶サブシステムを含むデータ処理システムの構成例を示す概念図である。

05 【図2】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置における共有メモリ(SM)内のキャッシュ管理領域とキャッシュメモリ部(CACHE)との関係を示した概念図である。

10 【図3】本発明の一実施の形態である記憶サブシステムの制御方法におけるキャッシュセグメントの新規確保時の処理の一例を示したフローチャートである。

【図4】本発明の一実施の形態である記憶サブシステムの制御方法におけるキャッシュセグメントの再確保要求時の処理の一例を示したフローチャートである。

15 【図5】本発明の一実施の形態である記憶サブシステムの制御方法におけるキャッシュメモリ部の利用率の関係の一例を示す概念図である。

20 【図6】本発明の一実施の形態である記憶サブシステムの制御方法におけるキャッシュメモリ部の利用率の関係の一例を示す概念図である。

【図7】本発明の一実施の形態である記憶サブシステムの制御方法におけるキャッシュメモリ部の利用率の関係の一例を示す概念図である。

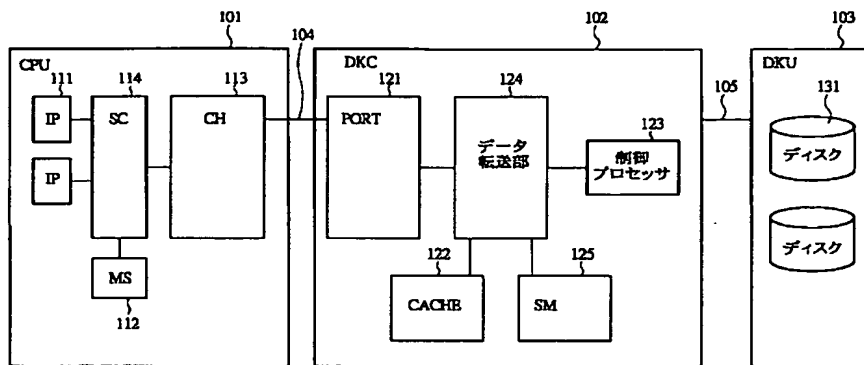
25 【図8】本発明の一実施の形態である記憶サブシステムの制御方法におけるキャッシュメモリ部の利用率の関係の一例を示す概念図である。

【符号の説明】

101…中央処理装置(CPU)、102…ディスク制御装置(DKC)(記憶制御装置)、103…ディスク駆動装置(DKU)、104…CPU-DKC間接続ケーブル、105…DKC-DKU間接続ケーブル、111…演算処理部(IP)、112…主記憶装置部(MS)、113…入出力チャネル部(CH)、114…システムコントローラ部(SC)、121…ポート部(PORT)、122…キャッシュメモリ部(CACHE)、123…制御プロセッサ部、124…データ転送制御部、125…共有メモリ(SM)、131…ディスク装置(記憶装置)、201…キャッシュ全セグメント数、202…使用セグメントカウンタ(第1記憶手段)、203…グループ単位セグメントカウンタ(第2記憶手段)、204～206…セグメント制御ブロック(SGCB)、214～216…セグメント使用状態、224～226…セグメントアドレス、230～237…セグメント。

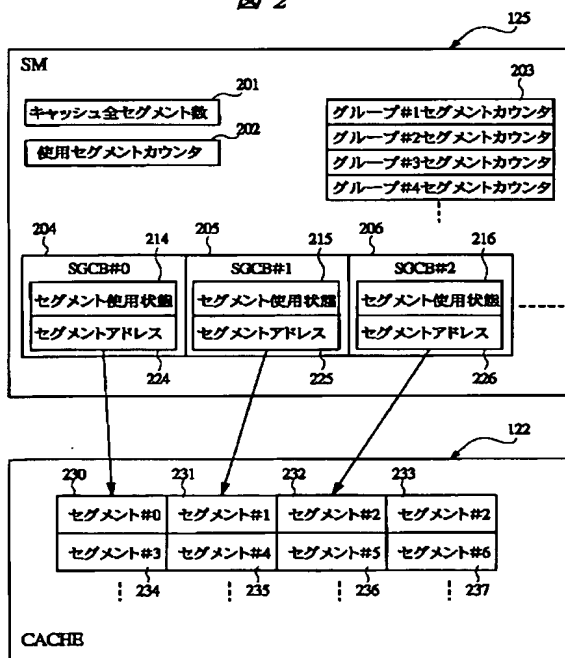
【図1】

図 1



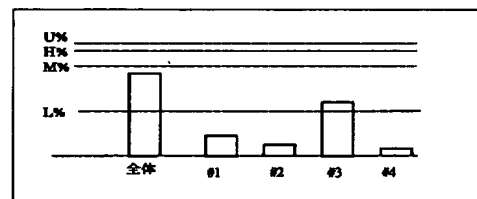
【図2】

図 2



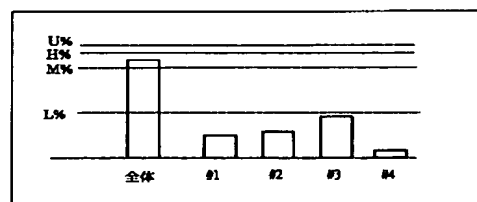
【図5】

図 5



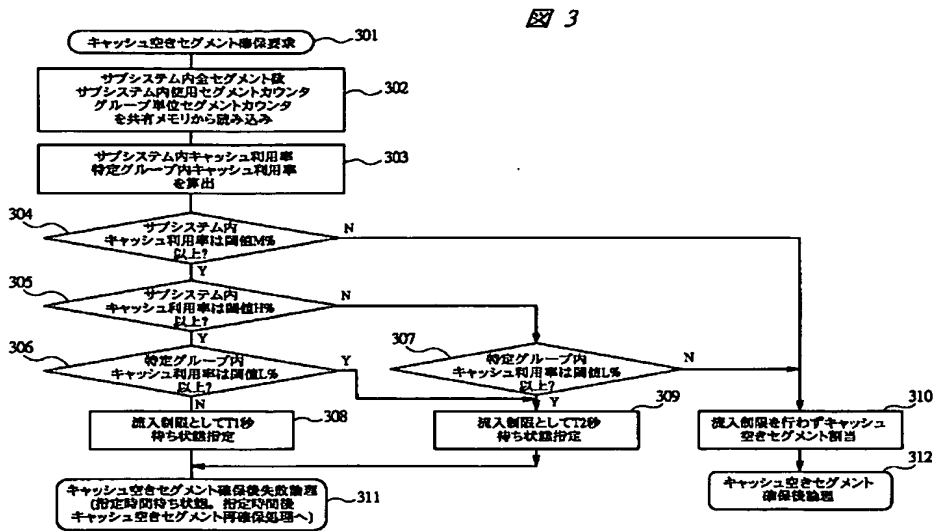
【図6】

図 6

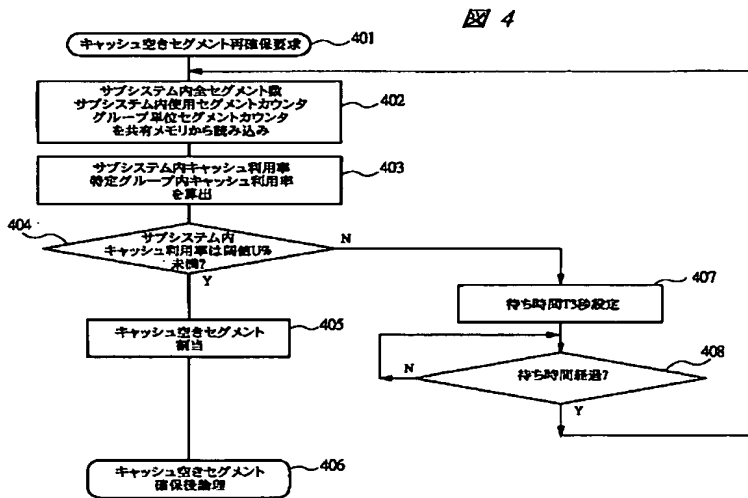




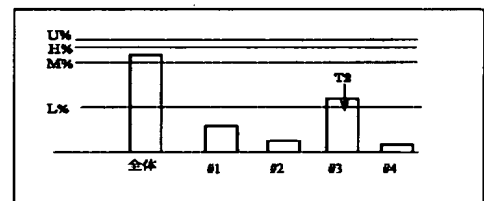
【図3】



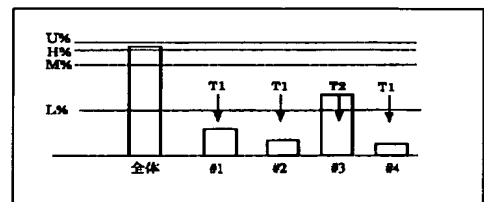
【図4】



【図7】



【図8】



フロントページの続き

(72)発明者 黒川 勇  
神奈川県小田原市中里322番地2号 株式  
会社日立製作所 R A I D システム事業部内 05

(72)発明者 川口 勝洋  
神奈川県小田原市中里322番地2号 株式  
会社日立製作所 R A I D システム事業部内  
Fターム(参考) 5B005 JJ13 MM11 QQ04 UU42 VV03  
5B065 BA01 CA12 CE12 CH02